A Time-Domain Derivation of Optimal and Suboptimal Kirchhoff Quantitative Migrations Via a Least-Squares Approach

Guy Chavent, René-Edouard Plessix

INSTITUT NATIONAL DE RECHERCHE EN INFORMATIQUE ET EN AUTOMATIQUE

A time-domain derivation of optimal and suboptimal Kirchhoff quantitative migrations via a least-squares approach

Guy Chavent and René-Edouard Plessix

N° RR-2967

September 4, 1996

THEME 4

Rapport de recherche

ISSN 0249-6399

A time-domain derivation of optimal and suboptimal Kirchhoff quantitative migrations via a least-squares approach

Guy Chavent *and René-Edouard Plessix †

Thème 4 — Simulation et optimisation
de systèmes complexes
Projet ESTIME

Rapport de recherche n° RR-2967 — September 4, 1996 — 38 pages

Abstract: We develop in this paper a general methodology for the derivation of optimal and suboptimal quantitative migration formula from the data misfit function associated to a given forward modeling operator. By construction, these migrations take into account any feature and approximation which has been incorporated into the data misfit function, such as finite aperture, band limited source, surface boundary conditions, multishot data, etc...

This methodology is then applied to the case where the forward modelling is made via the Born plus rays approximation and the reflectivity of the earth is represented by an array of diffracting points. This leads to the construction of efficient suboptimal quantitative Kirchhoff migration formula which provide a good restitution of amplitudes under a wide variety of circumstances (finite aperture, band limited source, etc...) and for a wide range of propagators (slowness background).

Key-words: migration, seismic inversion

(Résumé : tsvp)

* CEREMADE, Université de Paris-Dauphine, pl. du Maréchal De Lattre De Tassigny, F-75775 Paris Cédex 16, and INRIA-Rocquencourt, Domaine de Voluceau, B.P. 105, F-78153 Le Chesnay Cédex.

† IFREMER, Centre de Brest, Structures et Ouvrages en Mer, B.P. 70, F-29280 Plouzané, and INRIA-Rocquencourt.

Unité de recherche INRIA Rocquencourt
Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
Téléphone : (33 1) 39 63 55 11 – Télécopie : (33 1) 39 63 53 30

Migrations quantitatives de Kirchhoff optimales et sous-optimales, dans le domaine temps, via une approche moindres carrés

Résumé : Nous présentons dans ce travail une approche générale pour l'obtention de formules de migration optimales et sous-optimales à partir de la fonction coût d'erreur sur les données associée à une modélisation donnée. Par construction, ces migrations prennent en compte toutes les propriétés et limitations qui ont été incorporées dans la modélisation et la fonction coût, comme par exemple l'extension finie du dispositif d'enregistrement, les sources à spectre limité, la condition aux limites en surface, les données multiples, etc...
Cette méthodologie est ensuite appliquée au cas d'une modélisation par l'approximation de Born plus rais, avec une reflectivité représentée par une grille de points diffractants. On construit ainsi des formules de migration quantitative de Kirchhoff sous-optimales, qui procurent une bonne restitution des amplitudes dans des circonstances très variées (extension finie du dispositif d'enregistrement, source de spectre limité, etc...) et pour une large gamme de propagateurs (vitesse de référence).

Mots-clé : migration, inversion sismique

1 Introduction

Relatively soon after the introduction of the least-squares approach to inversion by Bamberger et al (1982), it was recognized by Lailly (1983) [12] and Tarantola(1984b) [15] that the gradient of the data misfit function they were used to calculate beared a strong ressemblance with respectively the wave equation reverse time migration and the Kirchhoff migration : the adjoint equation of the control theory was nothing but the back propagation of the residual, and both the gradient and the migrated section were obtained by (slightly different) correllation of the forward and backward propagated fields. This recognition was the basis of the saying "migration is the first step of inversion" (cf [15]).It opened the way to new usages of the wave equation migration, in particular in iterative migration algorithm, which could then be understood as minimization algorithms.
Surprisingly, the least-squares approach to the Kirchhoff migration has never been exploited systematically. The reason may be cultural: since the dawn of Geophysics, the Kirchhoff migration has been developped as an attempt to solve an integral equation relating the diffracted field at a point interior to the earth to the diffracted field at the surface of the earth, and hence to the data. We mention in references a few papers in this line we have been looking at, but they represent only an infinitesimal part of the geophysical literature on this subject. Approximately at the same time where migration was linked to the gradient of the data misfit function, Beylkin (1984 , 1985) recognized the strong ties that the Kirchhoff migration had with the inversion of a generalized Radon transform. This allowed a precise mathematical analysis of the resolution capacity of the Kirchhoff migration, but ignored the least-squares connection. This integral equation approach has produced popular "inversion formula" aimed at restoring the true amplitude of the reflectors in the migrated sections, like the ones of Docherty (1991), Keho and Beydoun (1988) and Bleistein (1987).
These inversion formula are established for a single shot and are optimal under the relatively strong hypothesis of infinite aperture and high frequency limit (Dirac source).

In this paper, we take full advantage of the links between migration and least-squares :

- in the first four sections, we develop a general methodology for the derivation of migration formula from the data misfit function $J(\pi, r)$ (here π denotes the propagator, for example the smooth part of the slowness, and r is a vector which describes the reflectivity of the medium): after recognizing the poor restitution of amplitudes provided by the simplest "gradient migration"

  $$m_0 = -\nabla_r J(\pi, 0),$$

  we state an "imaging principle", which defines a "descent migration" m by

  $$m = K m_0 = -K \nabla_r J(\pi, 0)$$

  where K is a diagonal matrix of positive weights (the vector m is a descent direction for J with respect to r at $r = 0$, hence its name). Of course the best choice for K would be the pseudo-inverse of the Hessian $H(\pi) = \frac{\partial^2}{\partial r^2} J(\pi, 0)$, as in that case the formula for

m mimicks a quasi-Newton step for the solution of the least-squares problem. But this is computationally unaffordable, so we indicate how to determine a weight K^* which provides the best diagonal approximation to the inverse of the Hessian. The corresponding migration $m^* = K^* m_0$ is called an optimal quantitative migration. Notice that, by construction, m^* will be optimal under any feature and approximation which has been incorporated in the definition of the data misfit $J(\pi, r)$, such as multishot data, band-limited source, finite aperture, surface boundary condition, etc...

- in the last four sections, we apply the above methodology to the case where the data misfit $J(\pi, r)$ is computed through a Born + rays approximation with the reflectivity of the medium represented by an array of diffracting points.
  We show that the corresponding class of descent migration formula, provided the proper weight K and the proper filtering and amplitude correction of the data are chosen, contains the Kirchhoff migrations. In particular, the above-mentionned "inversion formula" for a single shot gather belong to the class of descent migrations for this Born + ray approximation. Then we determine the optimal weight K^* in that class, and a computationnally less expensive suboptimal weight $\tilde{K}^*$. The corresponding optimal and suboptimal quantitative Kirchhoff migrations m^* and $\tilde{m}^*$ are hence expected to give, by construction, a better restitution of the amplitudes than the classical "inversion formula", in particular for multiple shot gather and finite aperture. Numerical results illustrate and confirm this expectation.

Throughout the paper, we have taken a great care to track down any simple explicit dependance of the weights and migrations on the propagator π (smooth slowness). This is very important in the applications we have in mind, where the migration operator is used inside a loop for the determination of π, as in migration velocity analysis for example. As a result of this care, our final suboptimal quantitative Kirchhoff migration formula is shown to provide a good resimulation of data over a large range of propagators around the one which has been used to compute the weights $\tilde{K}^*$.

2 Data misfit and gradient - migration

It will be convenient to describe the earth to be analysed by its propagator vector π and its reflectivity vector r . Vector π parametrizes the propagation of energy (for an acoustic model, π usually defines the smooth part of the slowness), and r the diffraction of energy (usually r defines the rough or small wavelength part of the earth material properties).
We refer to [7] for a general discussion of the propagator/reflectivity representation, and to formula (27) below for the case of a Born + ray forward modeling. Once a forward modelling procedure has been chosen, we shall denote by $c_{\pi,r}$ the collection of synthetic data $c_{\pi,r}^{S,G}(t)$ generated at geophone G for shot S, and by d the collection of corresponding recorded data $d^{S,G}(t)$.

The formula for the computation of the synthetic $c_{\pi,r}^{S,G}(t)$ under the Born + ray approximation are given in (37)(32) below and established in Appendix 1. The classical data misfit function associated to the (π, r) earth parameters is then

$$J(\pi,r) = \frac{1}{2}\|d - c_{\pi,r}\|^2 = \frac{1}{2}\sum_S \sum_G \int_0^T (d^{S,G}(t) - c_{\pi,r}^{S,G}(t))^2 dt. \tag{1}$$

It is understood here that d and $c_{\pi,r}$ represent the data and synthetics after any preliminary treatment (deconvolution, compensation for geometrical spreading... as in (37) below for example) has been applied.

It is well known that the objective function $J(\pi, r)$ is of little practical use for the waveform inversion of the data d, in particular because of its multiple local minima with respect to the propagator π. However, for a fixed propagator π, $J(\pi, r)$ has a nice behaviour as a function of the reflectivity r: it is almost quadratical, and departs from quadraticity only because of the presence of multiple reflections. Moreover, it is well known, since [12] and [14], that its gradient with respect to r is a migrated section of the data:

Definition 1. *The operator $\mathcal{M}_0 : d \rightarrow m_0$ from the data to the reflectivity space is a gradient-migration operator if and only if:*

$$m_0 = -\nabla_r J(\pi, 0) \tag{2}$$

Here m_0 is called the gradient-migrated section of the data d with the propagator π.

This concept of gradient-migration has not been very successful until now for the design of migrations from a practical point of view; the reason for this is that it usually gives a very poor restitution of the amplitudes of the migrated events (deep events tend to be strongly attenuated). So gradient-migration were easily beaten by the so-called "inversion formula" of [11] [5] for exemple, which were based on a careful inversion, in the high frequency limit, of the Generalized Radon Transform appearing in the linearized integral equation for the diffracted field, and gave a much better restitution of the amplitudes of migrated events.

However, compared to this classical approach, the gradient-migration has the advantage of flexibility: by construction, it is based solely on the data misfit function $J(\pi, r)$, and so is able to take into account any information which has been incorporated into $J(\pi, r)$ through the modelling step:

- variable background π
- surface boundary condition
- finite number of shots and their locations
- for each shot, finite number of receivers and their locations

- band-limited wavelet (if it is known !)
- preliminary deconvolution and amplitudes correction of the data
- etc...

Some of these features, like for example the aperture effects created by the finite number of shots and receivers, are very difficult, if not impossible, to take into account in the classical approach.

Our objective in this first part of the paper is to define a methodology for the construction of quantitative migrations, which retains the advantages of both approaches: flexibility of gradient-migration, and a good restitution of the amplitudes of migrated events.

We devote the three next paragraphs to the description of this methodology:

- the first step is the introduction of the concept of descent - migration, which is a slight generalization of that of gradient-migration; it produces a large class of migration operators, which retains the flexibility of gradient-migrations, but also contains as special cases the "inversion formula" of [11] and [5].
- in the second step, we shall propose a guideline for the selection, in this large class, of one descent-migrations which provides a good restitution of the migrated events, by requiring that it tries to mimick one Newton step for the inversion of $J(\pi, r)$ with respect to r starting from $r = 0$. Such a migration will be called quantitative.
- in the third step, we shall indicate how to select, in the class of descent-migrations, the one which mimicks at best (in a sense we shall make precise) the Newton step: such a migration will be called an optimal quantitative migration.

3 Descent migrations

Let us denote by R the space of reflectivity vectors r. By construction, r parametrizes the rough part - or small scale variations - of the earth parameters. Hence to any vector m of R we can associate the image, in the (horizontal distance) $\times$ (depth) domain, of the rough part of the earth parameters. So we introduce a new imaging principle, based on the data misfit function J only:

Definition 2. *(Descent imaging principle) Given a data misfit function $J(\pi, r)$, a reflectivity vector m of R is a migrated image of the seismic data d if and only if it is a descent direction for $J(\pi, r)$ with respect to r at $r = 0$, i.e.:*

$$J(\pi, 0 + \lambda m) < J(\pi, 0) \quad \text{for } \lambda \text{ small enough} \tag{3}$$

or equivalently

$$< m, \nabla_r J(\pi, 0) >_R \quad < 0 \tag{4}$$

or

$$< m, m_0 >_R \quad > 0 \tag{5}$$

where $<,>_R$ *represent the scalar product in the reflectivity space* R *and* m_0 *is the gradient-migrated section of definition 1.*

The motivation for this definition is clear: if m satisfies (3), this means that λm generate a synthetic c which contains events which will substract partially, in (1), from d in order to decrease J: this means that the reflectors in m are located at the right place to generate some of the events of d: but there is no guarantee concerning the amplitudes (absolute and relative) of the events in m.

Of course, this imaging principle is quite weak, and there are an infinity of migrated sections m of a given time section d which satisfy (3), (4) or (5) ! In particular, the gradient migrated section m_0 of definition 1 is a migrated image in the sense of this imaging principle, as

$$< m_0, m_0 >_R = \|m_0\|^2 > 0.$$

The most natural way to obtain a migrated image using this imaging principe is then to premultiply the gradient-migrated section m_0 by a symmetric positive matrix K, as:

$$< Km_0, m_0 >_R \quad \geq 0 \tag{6}$$

Comparing (5) and (6) we see that $m = Km_0$ will be a migrated section as soon as Km_0 is not orthogonal to m_0, which will be the case for all reasonnable choices of K.
The matrix K, which maps the reflectivity space R into itself, is extremely large. In order to limit the amount of work required in first choosing K and second computing Km_0, we shall restrict ourself to the consideration of diagonal matrices K. This leads us to the

Definition 3. *the operator* $\mathcal{M} : d \rightarrow m$ *from the data to the reflectivity space is a descent-migration associated to the data misfit function* $J(\pi, r)$ *and to the diagonal positive migration weigh matrix* K *if:*

$$m = Km_0 \quad \textit{with} \quad m_0 = -\nabla_r J(\pi, 0) \tag{7}$$

Here m *is called the descent-migrated section of the data* d *with the propagator* π.

Hence, once the data misfit function $J(\pi, r)$ has been defined by taking into account all the required feature of the problem in the forward modelling, there are as many descent-migrations as positive diagonal matrices K on the reflectivity space ! We take advantage in the next paragraph of this degree of freedom to enhance the restitution of the amplitudes of the migrated events.

4 Inversion as a guide to quantitative migration

For a given propagator π, the best possible migrated image of the data d is the (least-squares) inverted section $\hat{r}$ which minimizes $J(\pi, r)$ over the reflectivity space R; it satisfies the first order necessary condition:

$$\nabla_r J(\pi, \hat{r}) = 0 \tag{8}$$

As J is almost quadratical in the respect to r, the minimizer $\hat{r}$ can be estimated, with a very good approximation (and no approximation at all when a linearized forward model is used as for example the Born + Ray approximation) by performing one Newton step starting from $r = 0$:

$$H(\pi)\hat{r} \simeq -\nabla_r J(\pi, 0) \tag{9}$$

where the Hessian

$$H(\pi) = \frac{\partial^2 J}{\partial r^2}(\pi, 0), \tag{10}$$

is a square matrix mapping the reflectivity space R into itself. As J is almost quadratical with respect to r, $H(\pi)$ will in practice be semi-definite positive. But it will not in general be invertible, as for example changing the reflectivity r in areas which are not illuminated by any shot will have no effect on the synthetic $c_{\pi,r}$: such perturbations are in the nullspace of $H(\pi)$, which hence contains non-zero vectors. So equation (9) has in general many solutions $\hat{r}$, and one has to specify the one one is interested in. In absence of any a priori information on $\hat{r}$, it is reasonnable to search for the inverted section $\hat{r}_L$ which, among all solutions of (9), has the smallest semi-norm $\|\hat{r}\|_L$ defined by:

$$\|r\|_L^2 = < Lr, r >_R, \tag{11}$$

where L is a diagonal matrix with entries larger than or equal to zero, and such that the nullspaces of $H(\pi)$ and L have only the zero vector in common. Of course, one can use $L =$ Identity, so that $\|r\|_L$ coincides with the usual norm $\|r\|$, but we shall see in the application to Kirchhoff migration that it can be interesting to use a semi-norm $\| \quad \|_L$ adapted to the Hessian $H(\pi)$. This minimum-seminorm solution $\hat{r}_L$ of (9) is given by a pseudoinverse $H(\pi)_L^\dagger$ of $H(\pi)$:

$$\hat{r}_L \simeq -H(\pi)_L^\dagger \nabla_r J(\pi, 0) \tag{12}$$

We shall use the formulation (12) of the Newton step for the estimation of the inverted reflectivity section $\hat{r}_L$ as a guide for the choice of the positive diagonal migration weight matrix K in the descent-migration m:

$$m = -K \nabla_r J(\pi, 0) \tag{13}$$

By comparing (12) and (13) we are led to the:

Definition 4. *The descent migration operator associated to data misfit function $J(\pi, r)$ and the diagonal positive migration weight matrix K is a quantitative migration if the weight matrix K is an approximation to the pseudoinverse $H(\pi)_L^\dagger$ of the hessian $H(\pi)$ of the objective function with respect to r at zero reflectivity.*

So, if we are given a noiseless data d generated, with propagator π, by a time reflectivity vector r_{true}, a quantitative migration of d will produce a migrated section m which is an approximation to the minimum semi-norm $\hat{r}_L$ inverted section which itself is very close, in the area illuminated by the shots, to the true reflectivity section r_{true}. So quantitative migrations will tend to restitute properly the amplitudes of migrated events, the quality of this restitution depending of course of the quality of the approximation of $H(\pi)_L^\dagger$ by the diagonal matrix K !

As we shall see in the application to Kirchhoff migration, the Hessian $H(\pi)$ tends to be strongly concentrated along its diagonal, this concentration increasing with the number of shots and the recording offset. Hence its pseudoinverse $H(\pi)_L^\dagger$ can be approximated by a diagonal matrix K with a relatively good precision.

We conclude this paragraph by giving a constructive way to recognize that a positive diagonal matrix K is an approximation to $H(\pi)_L^\dagger$. Given a regularization parameter $\varepsilon > 0$, we first replace equation (9) (which has more than one solution) by its $L-$ regularized version:

$$(H(\pi) + \varepsilon L)\hat{r}_{L,\varepsilon} = -\nabla_r J(\pi, 0) \tag{14}$$

which has a unique solution as, by hypothesis, the nullspaces of $H(\pi)$ and L have only the zero vector in common. Then we known, from the theory of regularization (cf [10]), that

$$\hat{r}_{L,\varepsilon} \to \hat{r}_L \text{ in } R \text{ where } \varepsilon \to 0 \tag{15}$$

This shows that $(H(\pi) + \varepsilon L)^{-1}$ is an approximation of the pseudoinverse $H(\pi)_L^\dagger$ by a symmetric positive definite matrix. If we now substitute (14) into (13) we find that:

$$m = K(H(\pi) + \varepsilon L)\hat{r}_{L,\varepsilon}. \tag{16}$$

So the descent-migrated section m will be close to $\hat{r}_{L,\varepsilon}$, and hence to $\hat{r}_L$, as soon as the positive diagonal weight matrix K is a good left inverse of the positive definite matrix $H(\pi) + \varepsilon L$:

$$K(H(\pi) + \varepsilon L) \simeq I \tag{17}$$

Satisfying (17) gives a constructive way of chosing the diagonal positive migration weights K in such a way that $m = -K\nabla_r J(\pi, 0)$ is a quantitative migration of the seismic data d.

5 An optimal quantitative migration

We suppose in this paragraph that we know the Hessian $H(\pi)$ - this will be the case for example in the application to the Kirchhoff migration given in the second part of the paper, where one has a simple closed form formula for the elements of $H(\pi)$ - and we search for the best migration weight matrix K in a sense we make now precise:

Definition 5. *Given the data misfit function $J(\pi, r)$ and its Hessian $H(\pi)$, $\varepsilon > 0$ and a diagonal positive matrix L, the diagonal migration weight K^* is optimal if and only if:*

$$\begin{cases} \|K^*(H(\pi) + \varepsilon L) - I\|_\infty \leq \|K(H(\pi) + \varepsilon L) - I\|_\infty \\ \textit{for all diagonal matrices } K \end{cases} \tag{18}$$

Here $\| \|_\infty$ is the matrix norm associated to the $\| \|_\infty$ vector norm:

$$\begin{cases} \|A\|_\infty = \sup_{\|x\|_\infty = 1} \|Ax\|_\infty \\ \|x\|_\infty = \max_{i=1-n} |x_i| \end{cases} \tag{19}$$

This norm has been chosen in order to ensure an easy solution of the problem (18), which is given by (see Appendix 3 for the proof):

Proposition 1. *Given $J(\pi, r)$, ε and L, define:*

$$\beta_i = \frac{\sum_{j \neq i} |H(\pi)_{i,j}|}{|H(\pi)_{i,i} + \varepsilon L_i|} \quad i = 1, 2, ..., n. \tag{20}$$

Then the optimal migration weight K^* is given by:

i) *at lines i where $\beta_i < 1$, ie where $H(\pi) + \varepsilon L$ is diagonal dominant:*

$$K_i^* = 1 \ / \ (H(\pi)_{i,i} + \varepsilon L_i) \tag{21}$$

ii) *at lines i where $\beta_i \geq 1$, ie where $H(\pi) + \varepsilon L$ is not diagonal dominant:*

$$K_i^* = 0. \tag{22}$$

Then:

$$m^* = -K^* \nabla_r J(\pi, 0) \tag{23}$$

is the sought optimal quantitatively migrated section, which satisfies :

$$\|m^* - \hat{r}_{L,\varepsilon}\|_\infty \leq \max_{i \ st. \beta_i < 1} \beta i \tag{24}$$

where $\hat{r}_{L,\varepsilon}$ is the regularized approximation, given by (14) of the minimum-L-norm inverted reflectivity $\hat{r}_L$.

The proposition (1) is the basis for the design of optimal migration weights. In many situation, the Hessian $H(\pi)$ alone will not be diagonal dominant at any line (this will be the case in the application to the Kirchhoff migration presented below). In that case, it will be necessary to use a non-zero regularization term εL. From (20) (24) one sees easily that, when the size ε of this regularization term increases, the error between m^* and $\hat{r}_{L,\varepsilon}$ decreases. But at the same time the error between $\hat{r}_{L,\varepsilon}$ and $\hat{r}_L$ increases as a consequence of (15). A good choice of the regularization term should tend to balance these two errors. As this is very difficult to achieve pratically, we simply present a regularization which has proven to work well in our numerical experiments:

$$\varepsilon = 1 \quad , L_i = \sum_{j \neq i} |H(\pi)_{i,j}| \tag{25}$$

which ensures that $H(\pi) + \varepsilon L$ is diagonal dominant at all lines, and leads to the optimal weight:

$$K_i^* = (H(\pi)_{i,i} + \sum_{j \neq i} |H(\pi)_{i,j}|)^{-1}. \tag{26}$$

Notice that this weight is simply the inverse of the mass lumped Hessian $H(\pi)$!

In the four remaining sections, we shall apply the above methodology to the construction of an optimal and suboptimal quantitative Kirchhoff migration, based on the objective function $J(\pi, r)$ associated to the Born + ray forward modelling, and give numerical results which illustrate the construction and its performances.

6 The data misfit function under the Born + rays approximation

We define now the forward modelling and the data misfit function which, according to the results of the previous paragraph, will be the starting point for the construction of optimal and suboptimal quantitative Kirchhoff migration.

We consider a medium with known constant density, taken equal to one for simplicity, and use a linearized acoustic forward model (Born approximation). The slowness ν is splitted into its smooth and rough parts:

$$\nu = \nu_s + \nu_r$$

according to the frequency content of the source (cf [7]), and parametrized by the propagator vector π and reflectivity vector r as follows:

$$\left\{ \begin{array}{ll} \nu_s = & \pi \\ \nu_r = & \sum_M r_M \delta(x - x_M), \end{array} \right. \tag{27}$$

where we have used, as usual, the smooth part of the slowness as the propagator unknown in charge of describing the kinematic of the energy, and where the reflectivity of the earth is represented by a family of scattering points M located at the nodes of a rectangular grid (see figure 1). Given π and r, we want to compute the synthetic data $c^{S,G}$ recorded at a geophone G for a shot at point source S (figure 1). So we linearize the acoustic equation around the smooth background $\pi = \nu_s$.
The incident field $u_I(x,t)$ generated by the source in the smooth medium satisfies:

$$\nu_s^2 \frac{\partial^2 u_I}{\partial t^2} - \Delta u_I = f(t)\delta(x - x_S), \tag{28}$$

where:

$$f(t) \quad \text{is the band-limited source function} \tag{29}$$

The scattered field $u_S(x,t)$ corresponding to the reflectivity r satisfies:

$$\nu_s^2 \frac{\partial^2 u_S}{\partial t^2} - \Delta u_S = -2\sum_M \nu_{s,M} r_M \frac{\partial^2 u_I}{\partial t^2}\delta(x - x_M), \tag{30}$$

where:

$$r_M = \quad \text{scattering amplitude of point scatterer } M. \tag{31}$$

Equations (28) and (30) (plus adequate initial and boundary conditions) determine the scattered field u_S which is recorded at the geophones G.

Finally, using the high frequency asymptotic approximation (paraxial rays) for the solution of (28) and (30) (cf Annexe 1 and [13]), the synthetic seismogram recorded at time t at geophone G during the shot S is given by:

$$c_{ta}^{S,G}(t) = \sum_M (\pi_M)^s r_M A_M^{S,G} S^{(2+s)}(t - \tau_M^{S,G}), \tag{32}$$

where:

- the lowerscript ta stands for "true amplitude", emphasizing the fact that this synthetic tries to imitate the true amplitude data.
- the exponent s of π_M takes the values

$$s = 1/2 \text{ in } 2-D, \quad s = 1 \text{ in } 3-D \tag{33}$$

  and has nothing to do with the subscript s used to denote smooth quantities !
- the factor $(\pi_M)^s = (\nu_{s,M})^s$ comes from the term ν_s in the right-hand side of (30) and from the fact that, in 2-D, the attenuation factor between S and M is proportional to $\nu_{s,M}^{-\frac{1}{2}}$.

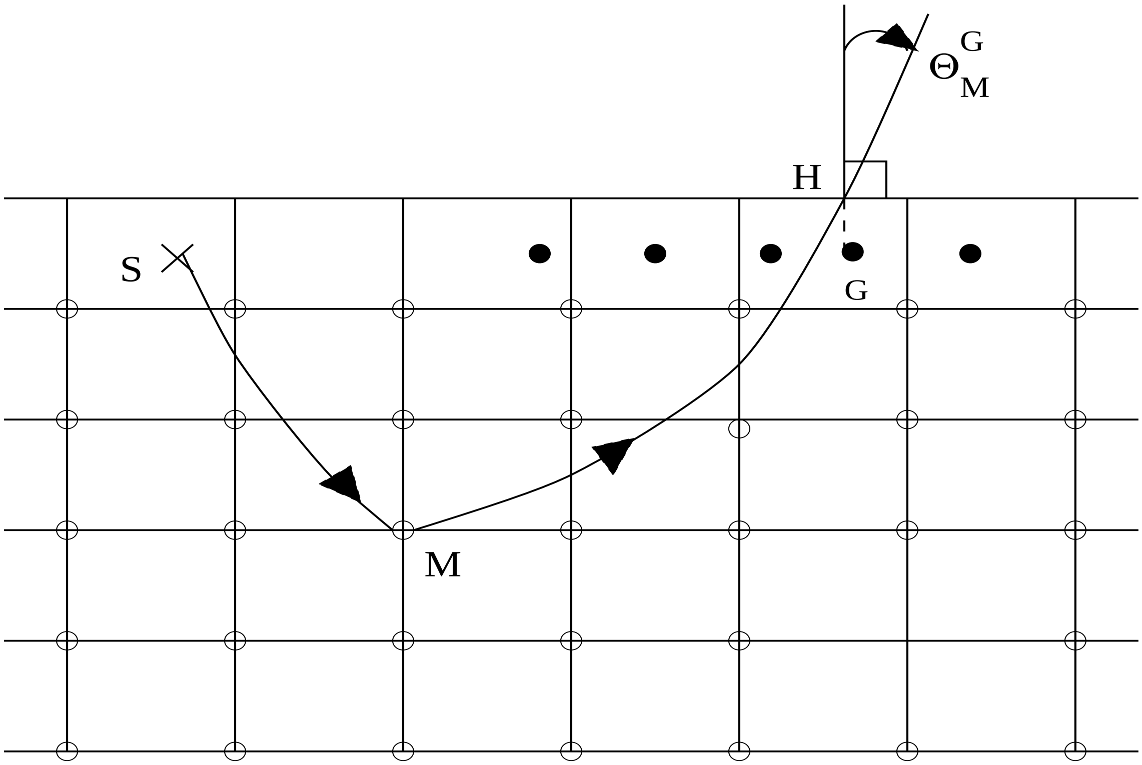

Figure 1: The geometry of the problem

- $A_M^{S,G}$ represent the part of the attenuation of S to G through M which depends on the propagator $\pi = \nu_s$ only through the travel time and geometry of the rays. Any explicit dependance on π_M has been factored out and transferred into $(\pi_M)^s$:

$$A_M^{S,G} = 4\Delta\tau^G \cos\theta_M^G B_M^{S,G} \tag{34}$$

where $\Delta\tau^G$ is the vertical travel time from geophone G to the free surface, θ_M^G is the emergence angle (see figure 1) and where $B_M^{S,G}$ is defined in Appendix 1 and depends only on the material properties at S and G and on the geometric divergence between S, M and M, G.

- $S(t)$ describes the shape of the wavelet signal radiated by the source, and which propagates in the smooth medium; it is given by:

$$S(t) = f(t) \text{ in } 3-D \quad , \quad S^{(1/2)}(t) = f(t) \text{ in } 2-D \tag{35}$$

and we shall suppose in the sequel that

$$\left\{ \begin{array}{l} S \text{ is three times continuously derivable, and} \\ \text{that } S(t) = 0 \text{ for } t < 0 \text{ and } t > T_S, \text{ where } T_S > 0 \text{ is the duration of the source} \end{array} \right. \tag{36}$$

- $\tau_M^{S,G}$ is the travel time from S to G through M.

Notice that, when the slowness smooth background ν_s is constant, then formula (32) is exact.

Suppose now that we are given a family of true amplitude shot-gather section $d_{ta} = (d_{ta}^{S,G})$ which we want to migrate. In order to compare this section with the synthetic section $c_{ta} = (c_{ta}^{S,G})$ defined by (32), we perform on both sections a preliminary treatment. We define for each source S and each geophone G:

$$d^{S,G}(t) = W_G(t) h * d_{ta}^{S,G}(t), \quad c^{S,G}(t) = W_G(t) h * c_{ta}^{S,G}(t) \tag{37}$$

where:

$$\left\{ \begin{array}{lcl} W_G(t) = t^s & = & \text{amplitude correction factor for the compensation of} \\ & & \text{the geometrical spreading} \\ h & = & \text{misfit filter (for a partial deconvolution of the data for example).} \end{array} \right. \tag{38}$$

We shall call $d^{S,G}(t)$ and $c^{S,G}(t)$ the misfit filtered data and synthetic. The seismic sections

$$d = (d^{S,G}), \quad c = (c^{S,G}) \tag{39}$$

are the one which enter in the definition (1) of the data misfit function $J(\pi, r)$, which is now perfectly defined.

7 Kirchhoff descent migrations

We recall first the classical approach to Kirchhoff migration: given a collection $d_{ta} = (d_{ta}^{S,G}(t)$, all $S, G, t)$ of seismic data, its Kirchoff migrated section m^K is given, at any point scatterer M, by (the upperscripts K stand for "Kirchhoff"):

$$m_M^K = \sum_S \sum_G (A^K)_M^{S,G} E^{S,G}(\tau_M^{S,G}) \tag{40}$$

where $E^{S,G}(t)$ is the true amplitude data filtered and amplified for migration:

$$E^{S,G} = W_G^K . g^K * (d_{ta}^{S,G})^{(s)} \tag{41}$$

with:

$$\begin{cases} W_G^K(t) & = \text{Kirchhoff amplitude correction factor} \\ g^K & = \text{Kirchhoff filter used to partially deconvolve } (d_{ta}^{S,G})^{(s)} \quad (\text{"spiking"}) \\ s & = 1/2 \text{ in } 2-D \text{ or/in } 3-D \quad (\text{as in } (33)) \end{cases} \tag{42}$$

and where $(A^K)_M^{S,G}$ is the attenuation factor associated to the $S \rightarrow M$ and $M \rightarrow G$ rays, which is specific of the chosen Kirchhoff migration formula.

The formula (41) corresponds to the current usage of the Kirchhoff migration, where the data are derived one half (in 2-D) or one (in 3-D) time, then deconvolved and corrected for geometrical spreading before being migrated.

The choice for the attenuation factors $(A^K)_M^{S,G}$ can range from the most simple one $((A^K)_M^{S,G} = 1!)$ when only the phase of the signal is taken into account, to more sophisticated formula aimed at restoring, in the migrated section m^K, the amplitudes of migrated events, in which case formula (40) is called a "true amplitude" migration or an "inversion formula". Such formulae have been derived, for the migration of a single shot-gather, by a careful inversion, in the high-frequency limit, of the integral equation satisfied by the diffracted field. For example, if we choose, in 3-D:

$$W_G^K(t) \equiv 1, \; (A^K)_M^{S,G} = 4\Delta\tau^G \cos\theta_M^G \frac{A_G^M}{A_M^S} \tag{43}$$

where A_M^S and A_G^M are the attenuation factors in the free-space Green functions, as given by (76) of Appendix 1, then (40) is a discrete version of the inversion formula (16) of [8], which itself is equivalent to equation (5) of [11], and to formula β of [5]:

$$m_M^K = \pi_M \sum_G 4\Delta\tau^G \cos\theta_M^G \left(\frac{D_M^S}{\pi_S D_G^M \pi_G}\right)^{\frac{1}{2}} (g^K * d_G^{ta})'(\tau_M^{S,G}) \tag{44}$$

where $\pi = \nu_s$ is the propagator and D_M^S and D_G^M the geometrical divergences. The classical "inversion formula" (44) is valid for the migration of single shotgather only, and is established under the hypothesis that the signal is recorded at all surface points and in the high frequency approximation (in other terms, the effect of finite aperture of the data and finite band width of the source are not taken into account).

We return now to the approach based on the data misfit function. We explicit for this the descent migrations associated, by definition 3, to the data misfit function (1) (37) (32) of the Born + rays approximation, and investigate their relations with the classical Kirchhoff formula (40) and (44).

We suppose from now on that recording time T, and the sets of diffracting points M, sources S and geophones G are chosen such that:

$$T \geq \max_{M,S,G} \{\tau_M^{SG}\} + T_S \tag{45}$$

where T_S is the duration of the propagating wavelet $S(t)$ (cf (35) (36)). This will ensure that for each source S we have at each geophone G a complete information on the waves diffracted by each scatterer M.

We suppose also that the weights $W_G(t)$ for the amplitude correction in the data misfit function satisfy:

$$W_G \text{ varies slowly compared to the signal } d_{ta}^{S,G} \tag{46}$$

Under these two hypothesis, the gradient of $J(\pi, r)$ with respect to r at $r = 0$ is given, at a given point scatterer M, by (cf Appendix 2)

$$[-\nabla_r J(\pi, 0)]_M = (\pi_M)^s \sum_S \sum_G A_M^{S,G} E^{S,G}(\tau_M^{S,G}) \tag{47}$$

where $E^{S,G}(t)$ is the migration-filtered data defined by:

$$E^{S,G} = W_G^2 h * h * S^{(2+s)} * d_{ta}^{S,G} \tag{48}$$

or equivalently

$$E^{S,G} = W_G^2 h * h * f^{(1+s)} * (d_{ta}^{S,G})^{(s)} \tag{49}$$

where h is the filter used in the data misfit (cf (37)), and where s is the exponent defined in (33).

Looking at the formula (47) for the gradient, we see that its right-hand side has the same structures as the Kirchhoff migration algorithm (40): the gradient of J at point scatterer M is obtained by migrating at M, with a proper attenuation $(\pi_M)^s$ A_M^{SG}, the migration filtered data E_G recorded at each geophone G at a time equal from the travel time: source $S \rightarrow$ scatterer $M \rightarrow$ geophone G, and by stacking all these quantities at each point M over the sources and receivers.

The relation between the Kirchhoff migration and the gradient of the objective function for the Born + ray approximation was recognized first by [14].

Let us now consider the descent-migration associated to the same data misfit and to a given diagonal positive migration weight K :

$$m = -K\nabla_r J(\pi, 0) \tag{50}$$

which is hence given, at each point scattered M, by:

$$m_M = K_M(\pi_M)^s \sum_S \sum_G A_M^{S,G} E^{S,G}(\tau_M^{S,G}) \tag{51}$$

with $E^{S,G}$ still given by (48) or (49). Comparing (51) with (40) shows that (51) is a Kirchhoff migration for the following choices of A^K, W^K and g^K:

$$(A^K)_M^{S,G} = K_M(\pi_M)^s A_M^{S,G} \tag{52}$$

$$W_G^K(t) = W_G(t)^2 \tag{53}$$

$$g^K = h * h * f^{(1+s)} \tag{54}$$

So we call (51) a Kirchhoff descent-migration, where the weight K_M at each scatterer M can be adjusted freely. The choice of K_M will be discussed in the next section.

We compare now the preliminary treatments (41) and (49) applied to the data to prepare them for migration in the classical and least-squares approaches.

Equation (53) suggests that the amplitude correction $W_G^K(t)$ to be used when preparing data for the Kirchhoff migration should be t^2 (for 3-D data) or t (for 2-D data).

Concerning the preliminary filtering of the data, (54) suggests to use a filter g^K related to the source function f. In the current practice of Kirchhoff migration g^K is chosen such that it performs a "spiking deconvolution' of the signal generated by reflectors. In order to explicit how this can be achieved by a proper choice of h, we first compute the shape of the signal generated in our model by a plane wave reflector; by the method of images, the reflected signal (i.e. the sum of all diffracted signals coming from all the point diffractors modelizing the reflector) has the same shape $S(t)$; hence the signal recorded at a given geophone G near the free surface has a shape $S^{(1)}(t)$, i.e. , using (35)

$$d_{ta}^{S,G}(t) \simeq \sum_{\text{reflectors } j} \alpha_j f^{(s)}(t) \tag{55}$$

Let us now choose the filter h in the data misfit such that it performs a "spiking deconvolution" of $(d_{ta}^{S,G})^{(q)}$, for some order of derivation q. From (55) this implies that:

$$h * f^{(s+q)} = \delta_B \tag{56}$$

where δ_B is a band-limited Dirac function.
Then the corresponding migration filter g^K is given by:

$$g^K = h * h * f^{(1+s)} = h^{(1-q)} * \delta_B. \tag{57}$$

The action of this filter on $(d_{ta}^{S,G})^{(s)}$ is given by :

$$g^K * f^{(2s)} = \delta_B * h^{(1-q)} * f^{(2s)} = \delta_B * h * f^{(2s+1-q)}. \tag{58}$$

Comparing with (56), we see that, if we chose $q = (1+s)/2$, one has :

$$g^k * f^{2s} = \delta_B * \delta_B,$$

which shows that g^K performs also a (less demanding) spiking deconvolution of $(d_{ta}^{S,G})^{(s)}$. We summarize this in the

Proposition 2. *The preparation of $E^{S,G}(t)$ from the true amplitude data $d_{ta}^{S,G}(t)$ by the classical receipes (41) (42) of the Kirchhoff migration can be mimicked in the least-squares approach by choosing in (37) an amplitude correction factor $W_G(t) = W_G^K(t)^{1/2}$ and a filter h which performs a partial deconvolution of $(d_{ta}^{S,G})^{((1+s)/2)}(t)$ - the latter equivalence holding stricto sensu only if the data are generated by plane reflectors.*

To conclude this paragraph we define

$$K_M^{IF} = \frac{1}{\pi_M (A_M^S)^2} = \frac{16\pi^2}{\pi_S} D_M^S \tag{59}$$

where the upperscript IF stands for "Inversion Formula", $\pi = 3.14$ and $\pi_S =$ the smooth slowness at source S... Then one has the

Proposition 3. . *The Kirchhoff descent migrations (51) reduces, in 3-D and for a single shotgather, for the choices $W_G(t) \equiv 1, h$ as in (56), and $K_M = K_M^{IF}$ defined in (59), to the inversion formula (44) of [8] [11] [5].*

We take now advantage, in the next section, of the degree of freedom K_M available in the definition of a Kirchhoff descent-migration to find the best possible one.

8 Optimal and suboptimal Kirchhoff quantitative migrations

We apply now to the Born + ray data misfit function $J(\pi, r)$ defined by (1) (37) (32) the proposition 1 which provides optimal choices for the migration weight K_M at each diffracting point M.

As, by construction, the Born + ray approximation is linear with respect to the reflectivity r, we can denote by $B(\pi)$ the corresponding forward model:

$$c_{\pi,r} = B(\pi)r. \tag{60}$$

Hence the Hessian of J with respect to r, as defined in (10), is of the form

$$H(\pi) = B(\pi)^T B(\pi). \tag{61}$$

Of course, $B(\pi)$ is a huge matrix of size ($\sharp$ of time data) $\times$ ($\sharp$ of point scatterers), whose column c_M associated to scatterer M is made of all the time sections generated by a reflectivity $r = \delta_M$ containing a single scatterer at point M with amplitude one.

The Hessian $H(\pi)$ is then a (less but still) huge matrix of size ($\sharp$ of point scatterers) $\times$ ($\sharp$ of point scatterers), whose column H_M associated to scatterer M is given by $H_M = B(\pi)^T B(\pi)\delta_M$, ie, using the $M-th$ column c_M of $B(\pi)$:

$$H_M = B(\pi)^T c_M \tag{62}$$

or

$$\left\{ \begin{array}{l} H_M = -\nabla_r J(\pi, 0) \\ \text{for the data } d = c_M \end{array} \right. \tag{63}$$

This gives a simple interpretation of the Hessian: its $M-th$ line or column is the gradient-migration of the seismic section generated by a unit diffracting point M. In other terms, the seismic acquisition process followed by a gradient-migration transforms a one-pixel reflectivity δ_M at M into the $M-th$ line or column H_M of the Hessian $H(\pi)$. As we shall see in the numerical results, H_M is (proportionnal to) a band-limited Dirac function centered at M: the "thickness" of this function gives the resolution power of the combined data acquisition + migration procedure, and the "amplitude" of this function will allow, as we shall see shortly, to determine an optimal migration weight K_M at point M.

It is easy, by a direct calculation, to obtain from equations (1) (37) (32) defining $J(\pi, r)$ a closed form formula for the element $H(\pi)_{M,P}$ of the Hessian associated to scatterers M and P:

$$H(\pi)_{M,P} = (\pi_M \pi_P)^s \sum_S \sum_G A_M^{S,G} A_P^{S,G} W_G(\tau_M^{S,G}) W_G(\tau_P^{S,G}) w * w(\tau_M^{S,G} - \tau_P^{S,G}) \tag{64}$$

where

$$w = h * S^{(2+s)}. \tag{65}$$

In order to apply proposition 1 to determine the optimal weights K_M^*, we investigate first the diagonal dominance, on a line or column M, of the Hessian $H(\pi)$. Because $w * w$ has always a maximum amplitude at 0, the maximum of the peak in the column H_M of the Hessian is always located at M, i.e. on the diagonal. Notice that this implies that reflectors will always be imaged, in descent-migrations, at zero-crossings of the migrated sections. Notice also that this is a good start for diagonal dominance: the element on the diagonal of $H(\pi)$ are always larger than the off-diagonal elements. However, this is not sufficient, and in fact $H(\pi)$ will never be diagonal dominant: because the size of the grid for the scattering points has to be chosen smaller than the spatial wavelenght, which determines the "tickness" of the peak in the column H_M, this peak will always contain, beside the point M itself, a couple (often 4) of neighbors of M with amplitudes of the same order of magnitude as M itself, which prevents the column H_M to be diagonal dominant.

So one has to resort to regularizaton to apply proposition 1. The choice (25) for the regularization parameter εL leads to an optimal migration weight K^*_M which is the inverse of the corresponding diagonal term of the mass-lumped Hessian $H(\pi)$:

$$K^*_M = (H(\pi)_{M,M} + \sum_{P \neq M} |H(\pi)_{M,P}|)^{-1} \tag{66}$$

It will appear in the numerical result section that the choice (66) for K_M leads to a good restitution (in the illuminated area of course !) of both relative and absolute amplitudes of migrated events. By construction - and this is one strength of the least-squares approach to migration - K^*_M takes into account the diffraction effects due to the finite length of the recording device, the variation in illumination caused by the variable number of shots which illuminate the point scatterers, the reflecting surface condition near the geophones, the finite bandwidth of the source function, the variable background π...
However, the preliminary determination of K^*_M by formula (66) is very time consuming, because of the summation, for each M, over the diffracting grid.

In order to define a computationally affordable suboptimal weight $\tilde{K}^*_M$, we factorize first in the formula (64) for $H(\pi)_{M,P}$, the part which depends strongly and directly on the current propagator π by writing:

$$H(\pi)_{M,P} = (\pi_M)^{2s} w * w(0) H_{red}(\pi)_{M,P} \tag{67}$$

where the reduced Hessian H_{red} is given by

$$H_{red}(\pi)_{M,P} = (\frac{\pi_P}{\pi_M})^s \sum_S \sum_G A^{S,G}_M A^{S,G}_P W_G(\tau^{S,G}_M) W_G(\tau^{S,G}_P) \frac{w * w(\tau^{S,G}_M - \tau^{S,G}_P)}{w * w(0)}. \tag{68}$$

The first and the last factor of (68) are close to one when P is close to M, and the remaining factors depend on π only through $A^{S,G}_M$ (ie through the emergence angle Θ^G_M and the geometric divergence $B^{S,G}_M$ as shown in (34)), and $W_G(\tau^{S,G}_M)$.

Then we replace the full mass-lumping of (66) by a partial one limited to the nodes of a neighborhood C_M of M, and chose the suboptimal weights $\tilde{K}^*_M$ proportional to the inverse of the diagonal of the partially lumped Hessian:

$$\tilde{K}^*_M = (\pi_M)^{-2s} \tilde{K}^*_{red,M}, \tag{69}$$

where

$$\tilde{K}^*_{red,M} = \lambda \{ \sum_{P \in C_M} |H_{red}(\pi)_{M,P}| \}^{-1}. \tag{70}$$

The suboptimal Kirchhoff quantitative migration corresponding to the choice $K_M = \tilde{K}^*_M$ in (51) is then:

$$\tilde{m}^*_M = (\pi_M)^{-s} \tilde{K}^*_{red,M} \sum_S \sum_G A^{S,G}_M E^{S,G}(\tau^{S,G}_M) \tag{71}$$

In order for the migration formula (71) to restore at best the absolute amplitudes of the migrated events, the scaling coefficient λ in (70) is evaluated as follows: given a collection of data d (for example the one to be inverted), λ is determined by requiring that the synthetic $c = B(\pi)\tilde{m}^*$ resimulated from d using the a-priori guess π of the propagator is as close as possible to the data d. As c is clearly proportional to λ, this amounts to search for the minimum of a quadratic function, which is given by:

$$\lambda = +(c, d)/\|c\|^2 \tag{72}$$

where c is the resimulated synthetic corresponding to $\lambda = 1$, and where $(\, , \,)$ is the scalar product in the data space.

We conclude this section by some comments on the way the suboptimal Kirchhoff quantitative migration depends on the current propagator $\pi = \nu_s$. This issue is very important in situations where the migration is used inside a loop for the determination of the propagator (as for example in migration velocity analysis), as one does not want to have to recompute the weights $\tilde{K}^*$ each time the propagator is changed. We have carefully factored out in this paper any direct dependance with respect to the propagator, so that in the final formula (69), (70) for the suboptimal migration weights $\tilde{K}^*_M$, the reduced weight $\tilde{K}^*_{red}$ depends on $\pi = \nu_s$ only through kinematic and geometric quantities (travel times and geometrical divergence): these quantities can be computed with the initial guess of the propagator, stored and kept constant while π is modified, whereas the amplitude factors $1/(\pi_M)^{2s}$ are evaluated with the current propagator.

Notice also that the classical inversion formula (44) gives a "inverted" section m^K_M proportionnel to π_M. This is to be compared to the suboptimal Kirchhoff quantitative-migration (71) which is proportionnal to $1/(\pi_M)^s$: this shows that the inversion formula (44) will not provide a good relative amplitude of scatterers when a background π with strongly varying slowness is used.

9 Some numerical results

In order to test the different assumptions and algorithms described in the previous sections, we will first compute one line of the Hessian of J for different numbers of shots. Then we shall compare the suboptimal and optimal weight matrix $\tilde{K}^*$ and K^*, show their differences with the weight matrix K^{IF} of the classical inversion formula, and illustrate the quality of the suboptimal Kirchhoff quantitative migration on a synthetic seismogram created with the formula (37), (32). Finally, we shall see the effect induced on the quality of the resimulation by the preliminary evaluation of $\tilde{K}^*_{red}$ at a first guess π_0 of the propagator different from the current propagator π..

The numerical applications are based on the Born + rays forward modelling described in a previous section. The travel times are computed by ray tracing (the ordinary differential equation is solved with a Runge-Kutta method) and the geometrical divergences are approximated with a technique of paraxial rays.
We consider the 2D case with no filtering of the data ($h \equiv 1$) and no amplitude correction ($W_G(t) \equiv 1$). The synthetic seismogram at geophone G during shot S is then (cf (37), (32) and Appendix 1):

$$c^{S,G}(t) = \sum_M \sqrt{\pi_M}\; r_M \; A_M^{S,G} \; S^{(3)}(t - \tau_M^{S,G})$$

with

$$A_M^{S,G} = \frac{\Delta\tau^G \;\; \cos\theta_M^{S,G}}{2\pi \;\pi_G \; \sqrt{D_M^S D_G^M}},$$

and the migration operator becomes (cf (51))

$$m_M = K_M \; \sqrt{\pi_M} \; \sum_S \sum_G A_M^{S,G} \; E^{S,G}(\tau_M^{S,G}),$$

where K_M will be either the optimal K_M^* given by (66) or a suboptimal $\tilde{K}_M^*$ given by (69), (70), or the weight K_M^{IF} of the classical inversion formula.
The determination of the optimal K_M^* requires the computation of the Hessian of J, which is very expensive in CPU time and memory. In order to be able to compute these weights, we use a small synthetic case. The survey space is composed with 1000 m of water and 3000 m of earth, the acquisition device contains 24 receivers with 50 m of receiver interval, the offset between the source and the first receiver is 200 m and the shot interval is 25 m. The immersion of the source and the acquisition device is 10 m. The reflectivity is discretized on a grid constituted with 25 m by 15 m rectangular cells . In this example, for each shot, we use a grid of 1400 m by 3000 m (56 by 200 points), which is exactly below the source and the acquisition device and follows their displacements during the survey (this is equivalent to consider a fixed global grid for all the shots and to use, for each shot, a mask in order to eliminate the influence of the distant scattering points). For the computation of K^* and $\tilde{K}^*$, we only compute the Hessian coefficients on a grid (7 by 10 points per shot) coarser than the reflectivity grid and estimate the missing Hessian coefficients by linear interpolation. In

the following examples, two propagators are used: a constant propagator with the velocity equal to 1500 m/s (the velocity of the water) and a linear propagator varying from 1500 m/s at the top of the reflectivity grid to 4000 m/s at the bottom. The function $w = S^{(3)}$, which appears in the Hessian formula, is a band-limited Dirac function which lasts 0.03 s.

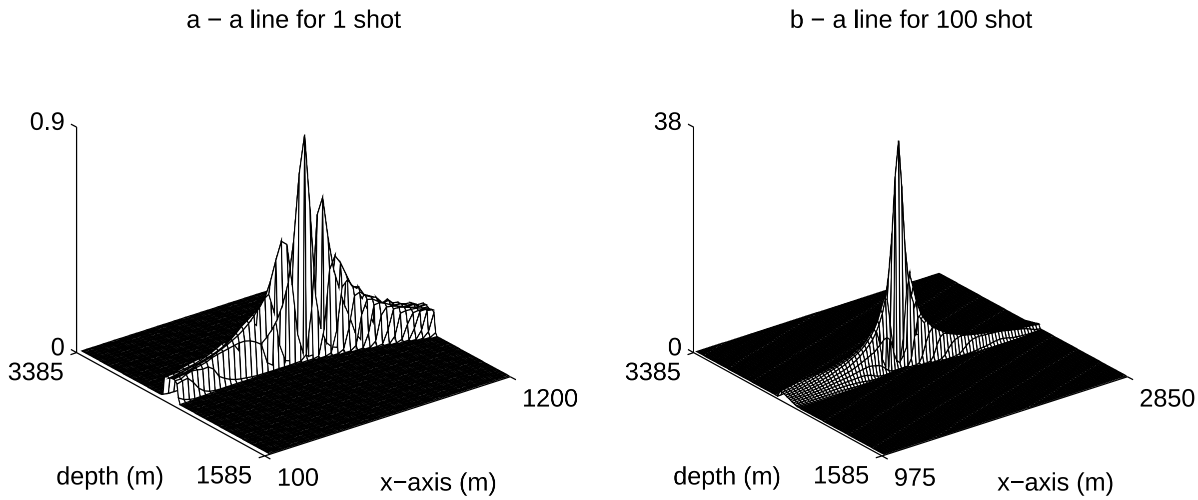

Figure 2: A line of Hessian matrix with a constant propagator (on the left, we consider only 1 shot, on the right, 100 shots)

The figure 2 represents (with the constant propagator) the line of the Hessian corresponding to a diffracting point located at the center of the reflectivity grid(as the elements of the line of the Hessian are indexed by the points diffractors of the reflectivity grid, this line is most naturally represented as a function on this grid): on the left, we consider only 1 shot (the first of the 100 shots), on the right, 100 shots. For 100 shots, the size of the reflectivity grid is 3850 m by 3000 m, however, for each shot, we only consider the part of this grid (1400 m by 3000 m) which is exactly below the source and the acquisition device (and then the other part of the grid has no influence). As it has been explained above, the Hessian has an important value on the diagonal but it is not a dominant diagonal matrix: in the two cases, the ratio $\sum_{P \neq M} |H_{MP}|/|H_{MM}|$ is bigger than 1, and decreases slowly when the number of shots increases (for 1 shot, it is equal to 50 and for 100 shots to 41). Of course, the peak is narrower with 100 shots (a similar behavior occurs when the number of receivers increases), but the number of non-zero terms becomes larger. Moreover, the terms around the diagonal point always have the same order of magnitude as the diagonal term itself. For these reasons, one needs to perform a certain amount of mass-lumping (ie to regularize) before estimating the weights (formula (66), (69), (70)).

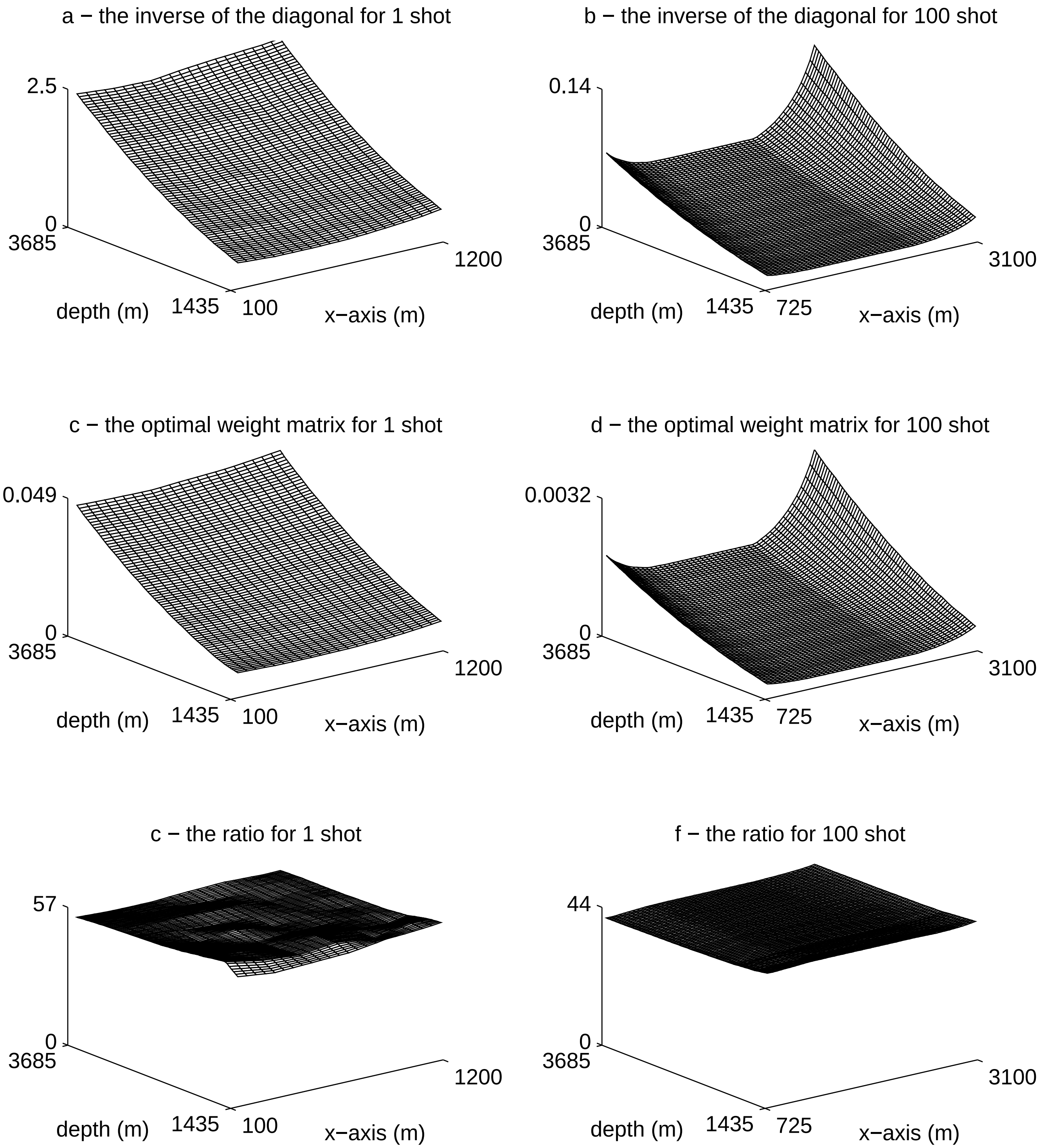

Figure 3: The suboptimal (top) and optimal weight (middle) matrix and their ratio (bottom) (on the left, we consider only 1 shot, on the right, 100 shots)

Therefore we now compare, with the constant propagator, the optimal weight matrix K^* given by (66)(the inverse of the diagonal matrix built by full mass lumping of the Hessian) and the suboptimal weight matrix $\tilde{K}^*/\lambda$ for $C_M = \{M\}$ (i.e. the inverse of the diagonal of the Hessian) (see formula (69) (70)). The figures 3.a, 3.b, 3.c and 3.d show these two weights for 1 and 100 shots (in order to obtain a significant graph, the parts of the graph close to the boundaries have been suppressed). We remark that their shapes are quite similar: it seems that K^* can be deduced from $\tilde{K}^*/\lambda$ by a scalar multiplication. To prove this assumption, we display, in the figures 3.e and 3.f, the division (term by term) of these two weights. Except near the top of the grid, they are almost proportional. The boundary effects come from the fact that the reflectivity grid has finite dimensions: near the boundaries, the set of the scattering points which have almost the same travel time as the diagonal point becomes smaller, so that the total mass decreases at such points faster than the peak value. Typically, this can be observed on the top of the grid, because the non-zero terms on a line of the Hessian principally correspond to scattering points above the diagonal point (see figure 2). In order to reduce these boundary effects, when the suboptimal weights are used in the migration, it is necessary to consider a reflectivity grid larger than the investigated area and to use only the part of the migrated section corresponding to this investigated area. In this way, we can replace the computation of the optimal weights K^* by the computation of the suboptimal weights $\tilde{K}^*$. This is numerically very important since the computation of the optimal matrix lasts 3 hours, for this example (on a Convex 3840 with one processor), whereas the computation of the suboptimal matrix lasts only 26 secondes.

With the linear propagator, the ratio $K^*/\tilde{K}^*$, with $\tilde{K}^*$ the inverse of the diagonal of the Hessian, is not quite constant (see figure 4.a). Therefore, the suboptimal weight matrix $\tilde{K}^*$ has to be improved in order to obtain a better approximation of the pseudo-inverse of the Hessian. This improvement is realized by enlarging the neighborhood C_M of M on which the partial mass lumping is performed. So instead of $C_M = \{M\}$, we use now for C_M a square of $(2i+1)^2$ points surounding M. We denote by $\tilde{K}_M^{(i)*}$ the corresponding suboptimal weights. A good choice would be to choose i such that $\tilde{K}^{(i)*}/K^*$ is almost constant over the diffracting grid. However, we don't want to compute the optimal weight K^*, since it is too expensive. So, we compute successively the weights $\tilde{K}^{(i)*}$ for $i = 0, 1, 2...$ and stop when the ratio $\tilde{K}^{(i+1)*}/\tilde{K}^{(i)*}$ becomes almost constant over the grid of diffracting points. In the example with the linear propagator, this led to the choice $i = 1$, and one can check on figure 4.b that the ratio $\tilde{K}^{(1)*}/K^*$ is almost constant. We notice that the size of the set C_M remains small, since the important values of the Hessian are close to diagonal points. Therefore, the over-cost to compute $\tilde{K}^{(i)*}$ is small. In the example, the computation of $\tilde{K}^{(0)*} = \tilde{K}^*$ lasts 26 secondes, the one of $\tilde{K}^{(1)*}$, 46 secondes, the one of $\tilde{K}^{(2)*}$, 72 secondes and the one of K^*, 3 hours. In order to see the differences with the classical inversion formula (44), which is optimal for one shot and for a full offset and frequence coverage (the stack with respect to the shots must be made afterward), we have shown K^{IF} in the figure 5.a (in 2D, $K_M^{IF} = 8\,\pi D_M^S$). We see that, without amplitude correction factor ($W_G(t) = 1$), this weight is different from the optimal or suboptimal (figure 3, left) ones. Indeed the matrix K^{IF} is only proportional to D_M^S (the geometrical divergence between the source and the scattering

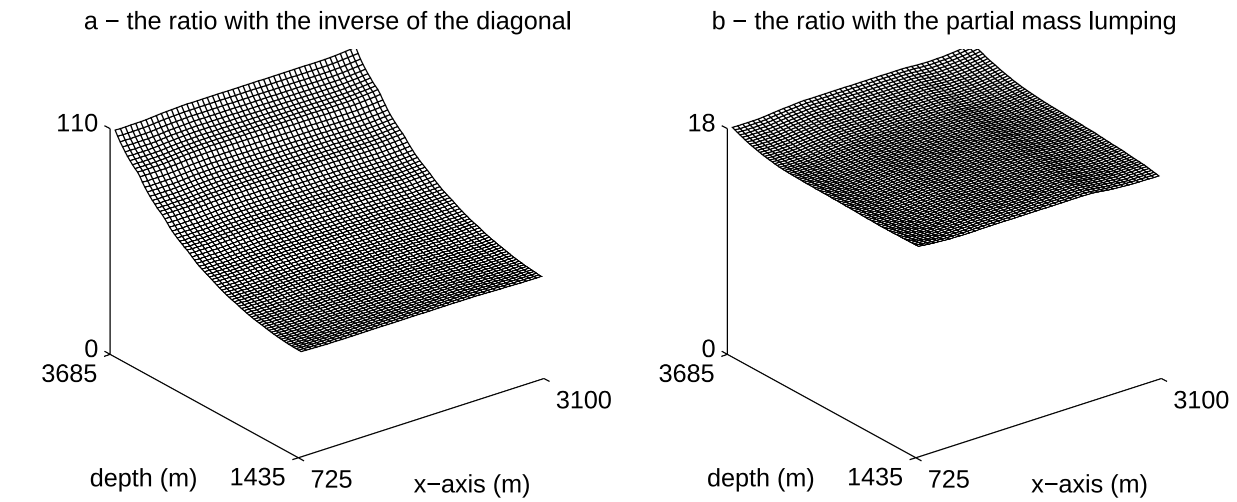

Figure 4: The ratio between the suboptimal weights and the optimal weights with the linear propagator: on the left $\tilde{K}^*/K^*$, on the right $\tilde{K}^{(1)*}/K^*$

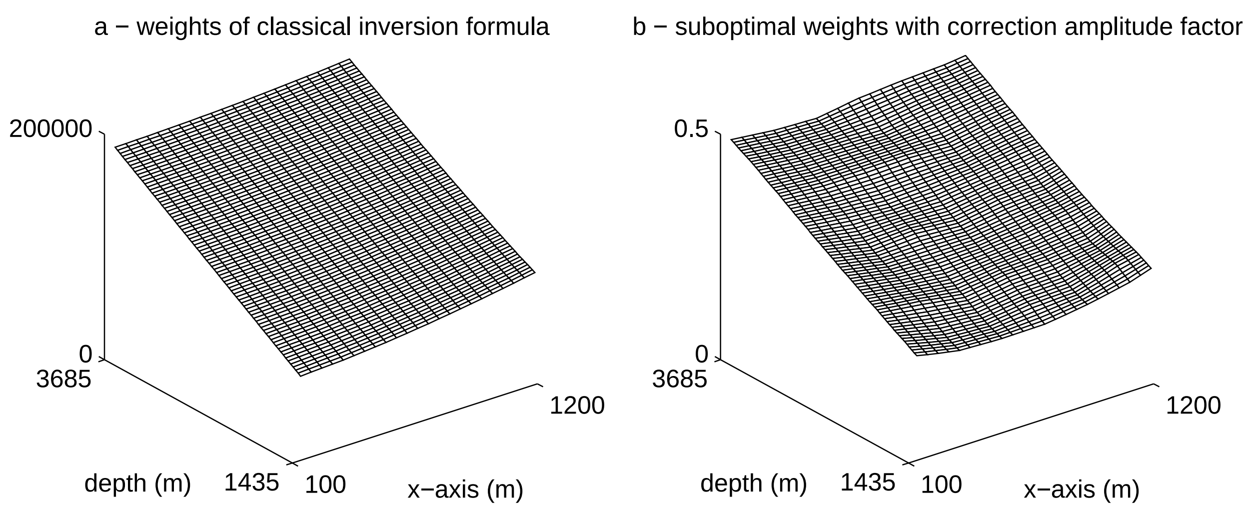

Figure 5: The weights of the classical inversion formula (on the left) and the suboptimal weights with an amplitude correction factor equal to $t^{1/2}$ (on the right)

point see (59)), whereas the suboptimal weights $\tilde{K}^*$ (inverse of the diagonal of the Hessian) is proportional to D_M^S and $(\sum_G (D_G^M)^{-1})^{-1}$ (D_G^M is the geometrical divergence between the scattering point and the receiver) (see (69), (70) and appendix A). Thereby, the deep reflectors are better migrated with the suboptimal weights, since they compensate for the attenuation which has occured during the backpropagation because of the finite aperture of the recording device.
In 2D, one can notice that K^{IF} on one side, and $\tilde{K}^*$ with $W_G = \sqrt{t}$ on the other side (see figure 5) have a similar behavior. However these two weights provide two different migrated sections since, with K^{IF}, no amplitude correction factor is used on the data (whereas, with $\tilde{K}^*$, this factor is $\sqrt{t}$).

In order to test the suboptimal Kirchhoff migration corresponding to the weights $\tilde{K}^{(1)*}$,

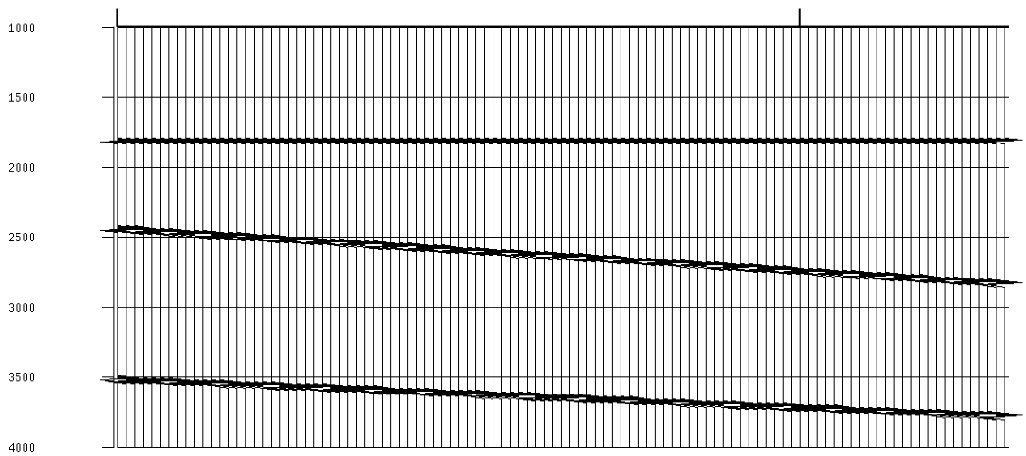

Figure 6: The true reflectivity

we have built synthetic data using the formula (32) with the reflectivity vector described in figure 6 (the 3 reflectors have the same amplitude) and the linear propagator. The figure 7 shows the result of the migration of the data for 100 shots with the suboptimal weights $\tilde{K}^{(1)*}$ with $\lambda = 1$. We notice that the reflectors are less sharp in the migrated section (see figure 8.a). To improve the approximation of the absolute amplitudes, we resimulate seismograms from this stacked, migrated section, and find the scaling factor $\lambda = \frac{(c,d)}{(c,c)} = 0.23$. With this coefficient, the ratio $\frac{J(\pi,\tilde{m}^*)}{J(\pi,0)} = 0.58$ ($\tilde{m}^*$ is the stacked, migrated section), which means that the resimulation explains 23 % of the data in the L^2 norm on the data space. We can compare, on the figure 8.a, the traces (located at 1500 m) of the true reflectivity and the migrated section and, on the figure 8.b, the traces 12 of the shot 50 of the data and the resimulated seismograms: the amplitudes are correctly restored (the differences between the peak values are likely due to the spreading of the events in the migrated section and the resimulated seismograms). The numerical results show that a descent migration with

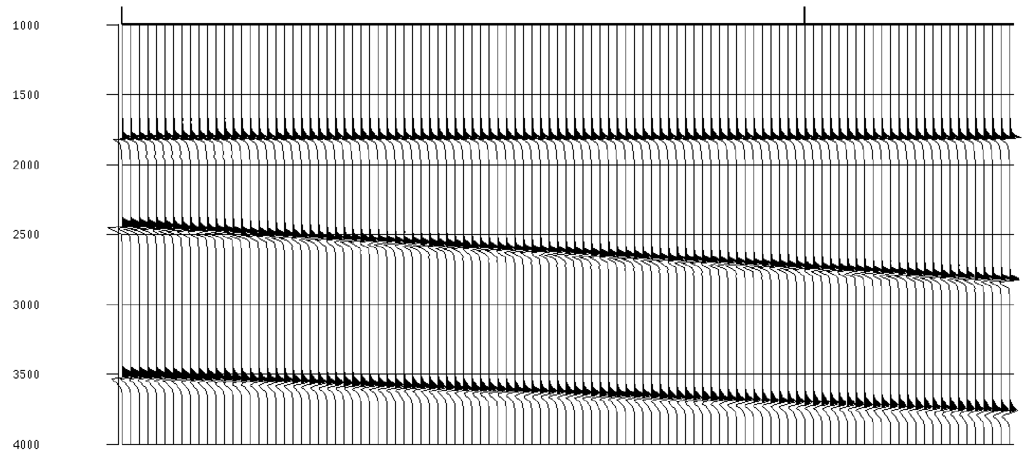

Figure 7: The stacked, migrated section

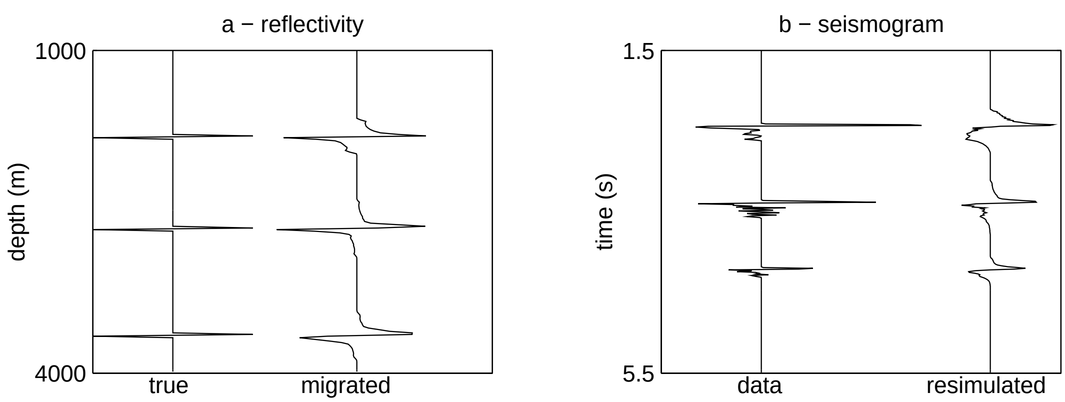

Figure 8: On the left, the comparison between a trace of the true reflectivity and the same one of the migrated section and on the right, the comparison between a trace of data and the same one of resimulated seismograms

the suboptimal weights provides a quantitative migration at a small CPU over-cost. This justifies the name of the suboptimal Kirchhoff quantitative migration we gave to formula (71).

It is interesting to visualize how the quality of the resimulation of a data d generated

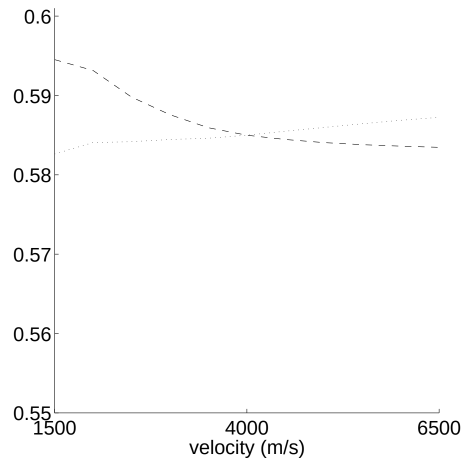

Figure 9: The squared misfit between the data and their resimulation as a function of the bottom velocity used to estimate the suboptimal weight (dashed line) or the suboptimal weights and the amplitude coefficients (dotted line).

by some true parameter π_t is affected when the suboptimal reduced weights and possibly also the amplitude coefficient are evaluated at some apriori guess π_0, different from π_t. So we have considered a range of a-priori guesses π_0, made of velocities obtained by linear interpolation between the surface velocity of 1500 m/s and a bottom velocity v chosen in the 1500-6500 m/s interval, and a data d generated with the ramp corresponding to bottom velocity $v = 4000$ m/s.

We display on figure 8 the squared misfit $J(\pi_t, r = m^*)$ between the data and its resimulation as a function of the bottom initial guess velocity v in two cases:

- dashed line: the migration m^* is performed with reduced weights evaluated with the initial guess π_0 (amplitude coefficients and, of course travel-times being evaluated at the true propagator π_t)
- dotted line: the migration m^* is performed with reduced weights and amplitude coefficients evaluated with the initial guess π_0 (travel-times being of course still evaluated at the true propagator π_t)

The travel-times are correctely computed in the migration steps, where the true propagator is used in both cases, so that the reflectors are correctely located in the stacked, migrated sections. Hence, the variations of J are only due to the variations in the amplitude of migrated and resimulated events due to the changes in reduced weights and amplitude coefficients. We notice, in figure 8, that they are small, less than 2%. This means that the resimulated seismogram remains close to the data, as soon as the travel times are well-known, even if the reduced weights of the quantitative migration and the amplitude coefficients are computed with a bad propagator.

Therefore, this allows to consider in a first approximation the reduced weights and amplitude coefficients as independent on the propagator for a velocity analysis without losing the quantitative character of our migration.

10 Conclusion

Using the least-squares approach to migration, we have defined a general methodology for the construction of optimal and suboptimal quantitative migrations, once the forward modelling and the norm in the data misfit have been chosen. By definition, these migrations are optimal under the same specification which are used in the construction of the forward model and data misfit, as for example source with finite band width, receivers with finite aperture, finite number of shots, deconvolution and equalization of data...

We have applied this methodology to a Born + rays forward modeling with a scattering approximation of the reflectivity, and to a data misfit defined on multishot finite aperture data, with preliminary filtering and amplitude correction. Our findings are:

- any classical Kirchhoff migration formula can be seen as a descent Kirchhoff migration for a properly defined data misfit function. Proposition 2 gives the precise relations which have to be satisfied in the preparation of the data in the two approaches if they are to be equivalent.

- A suboptimal quantitative Kirchhoff migration can be obtained by multiplying the gradient migrated section $m_{0,M}$ at scatterer M by a weight $\tilde{K}_M^*$ obtained by inverting the diagonal of the partially-lumped Hessian. The computation time required for the calculation of these weights is small. Our numerical results show that these suboptimal weights allow the same quality of restitution of the amplitudes in the migrated section as the fully optimal weights, which are computationaly much more expensive.

- The introduction of reduced weights $\tilde{K}^*_{red,M}$, which depend mainly on the geometry of shots, sources and receivers but little on the propagator π and can be evaluated preliminary using a first rough guess π_0 of π, allows to use the suboptimal quantitative Kirchhoff migration over a large range of propagators while maintaining the quality of the amplitude restitution and of resimulation.

Acknowledgments

The first author is indebted to Sam GRAY (Amoco Research Center, Tulsa) who helped him to find a path in the jungle of the literature on migration, and explained the grass-root nature of the Kirchhoff migration, to Francis COLLINO (INRIA, Rocquencourt) for fruitful discussions on the Born + rays approximation and to Eliane BECACHE (INRIA, Rocquencourt) whose careful reading of the paper led to the correction of an error. He would also like to acknowledge the hospitality of the Centre de Recherches Mathématiques from the Université de Montréal, where he was visiting while rewriting this paper. The second author wishes to thank Mr Y. H. De Roeck (IFREMER, France) for his support and his help.

A Asymptotic solution of the Born approximation

We recall here the main steps of the approximation of the linearized acoustic equations (28) and (30) by ray theory, as described for example in [4] and [13], which leads to the formula (32) for the synthetic seismograms.

Given the background propagator $\pi = \nu_s$, one first calculates (by ray tracing or finite difference solution of the eikonal equation for example):

$$\left\{\begin{array}{l} \text{For every source } S, \text{every grid point } M \text{ and every geophone } G \text{ (notation of figure 1) :} \\ \tau_M^S = \text{ travel time from source } S \text{ to } M \\ \tau_G^M = \text{ travel time from } M \text{ to } H \text{ (not } G!) \\ \theta_M^G = \text{ emergence angle at } H \text{ (not } G!) \text{ of the ray coming from } M \end{array}\right. \tag{73}$$

and:

$$\left\{\begin{array}{l} \text{For every geophone } G \\ \Delta\tau^G = \text{ vertical travel time from geophone } G \text{ to the free surface} \end{array}\right. \tag{74}$$

then:

- the incident field u_I, solution of (28), is approximated, at any point M by

$$u_I(M,t) = A_M^S S(t - \tau_M^S) \tag{75}$$

  where $S(t)$ is the propagating wavelet, defined in (35), and where A_M^S is the attenuation factor for the propagation from S to M, given, for a constant density $\rho = 1$, by

$$A_M^S = \frac{(\pi_S)^{s-1/2}}{2(2\pi)^s (\pi_M)^{1/2} (D_M^S)^{1/2}} \tag{76}$$

  where the exponent s is defined in (33), $\pi = 3,1416..., \pi_S$ and π_M denote the value of the propagator $\pi = \nu_s$ at the source and diffracting point location and where D_M^S

denotes the geometrical divergence along the ray going from S to M which can be computed by the technique of paraxial rays, or roughly approximated by its value in a constant velocity medium, given by:

$$D_M^S = |S - M|^{2s} \tag{77}$$

- the scattered field u_s, solution of (30), is approximated, at the point H of the free surface just above the geophone G, using the same approximation, by:

$$u_S(H,t) = -2\sum_M A_G^M \pi_M r_M D_t^{(1+s)} u_I(M, t - \tau_G^M) \tag{78}$$

  where the attenuation factor A_G^M along the ray going from the scatterer M to the surface point H above the receiver G is given by formula similar to (76) and (77), and where π_M is the value at scatterer M of the propagator $\pi = \nu_s$.

- the pressure field $c_{ta}^{S,G}(t)$ at geophone G can be computed using the fact that, when the scattering points M are far enough from the geophone, the scattered field u_S given by (78) is a superposition of (almost) plane waves, and that u_S is zero on the free surface just above G. So we can split (cf figure A-1) the wave field in the vicinity of the geophone G into the sum of an incident plane wave u_s and a reflected plane wave u_R:

$$\begin{aligned}
u_S(M,t) &= S(t - \frac{OM.n_S}{c}) && \|n_S\| = 1 \\
u_R(M,t) &= -S(t - \frac{OH.n_S + HM.n_R}{c}) && \|n_R\| = 1 \\
u(M,t) &= u_S + u_R = S(t - \frac{OM.n_S}{c}) - S(t - \frac{OH.n_S + HM.n_R}{c})
\end{aligned}$$

If the incident and reflected wave satisfy:

$$n_S = n_T + n_N, \; n_R = n_T - n_N,$$

then for all M of the surface, i.e. such that $HM.n_N = 0$, one has :

$$u(M,t) = S(t - \frac{OH.n_S + HM.n_T}{c}) - S(t - \frac{OH.n_S + HM.n_T}{c}) = 0$$

Hence $u(M,t)$ is the solution near the free-surface boundary. At a geophone $G = (x_G, y_G, z_G)$, the signal is:

$$c_G(t) = u(G,t)$$

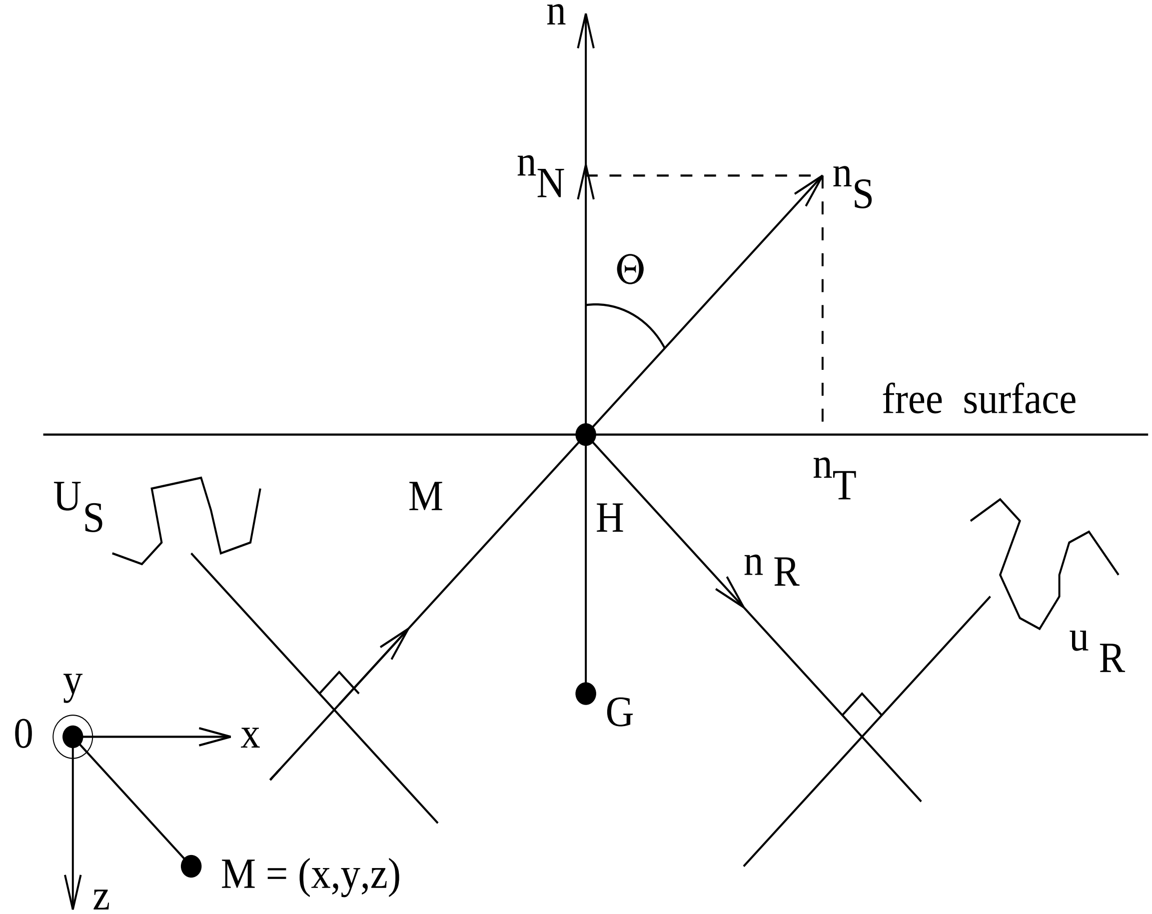

Figure 10: Notations for the calculation of the pressure field at a geophone G.

If the geophone depth z_G is a fraction of the spatial wavelength of the signal, one can write

$$\frac{\partial u}{\partial z}(H,t) \simeq \frac{u(G,t) - u(H,t)}{z_G - z_H}$$

so that, as $u(H,t) \equiv 0$:

$$c_G(t) \simeq (z_G - z_H)\frac{\partial u}{\partial z}(H,t)$$

But:

$$\frac{\partial u}{\partial z}(M,t) = -\frac{n_{S,z}}{c}S'(t - \frac{t - OM.n_I}{c}) + \frac{n_{R,z}}{c}(t - \frac{OH.n_S + HM.n_R}{c})$$

Hence, as $n_{S,z} = \cos\theta_M^G = -n_{R,z}$ and $n_{S,x} = n_{R,x} = n_x; n_{S,y} = n_{R,y} = n_y$

$$\frac{\partial u}{\partial z}(H,t) = -\frac{2\cos\theta_M^G}{c}S'(t - \frac{OH.n_I}{c}) = -\frac{2\cos\theta_M^G}{c}\frac{\partial u_S}{\partial t}(H,t)$$

An approximate expression for the signal at geophone G is hence

$$c_{ta}^{S,G}(t) \simeq -2\cos\theta_M^G \Delta\tau^G \frac{\partial u_S}{\partial t}(H,t) \tag{79}$$

where the lowerscript "*ta*" stands for "true amplitude", emphasizing the fact that this synthetic tries to restitute the amplitudes of data, and where:

$$\Delta\tau^G = \frac{z_G - z_H}{c} = \text{travel time from geophone } G \text{ to free surface.}$$

Remark 1. *Notice that, in the above formula, the emergence angle θ and the time delay $\frac{OH.n_S}{c}$ are evaluated at the point H of the free surface located above geophone G, and not at G itself.*

If we substitute (75) into (78) and (78) into (79) we obtain finally:

$$c_{ta}^{S,G}(t) = \sum_M 4\Delta\tau^G \cos\theta_M^G r_M \pi_M A_M^S A_G^M S^{(2+s)}(t - \tau_M^S - \tau_G^M) \tag{80}$$

But one can rewrite

$$\pi_M A_M^S A_G^M = (\pi_m)^s B_M^{S,G}$$

with the exponent s defined in (33), and $B_M^{S,G}$ defined by:

$$B_M^{S,G} = \frac{1}{4(2\pi)^{2s}} \times \frac{(\pi_S)^{s-1/2}}{(\pi_G)^{1/2}} \times \frac{1}{(D_M^S D_G^M)^{1/2}} \tag{81}$$

Now the synthetic rewrites:

$$c_{ta}^{S,G}(t) = \sum_M (\pi_M)^s r_M 4\Delta\tau^G \cos\theta_M^G B_M^{S,G} S^{(2+s)}(t - \tau_M^S - \tau_G^M) \tag{82}$$

which is the sought formula (32) with the notations (34). Notice that in (82), all the explicit dependance of the synthetic $c_{ta}^{S,G}$ on the propagator $\pi = \nu_s$ has been factored out in the term $(\pi_M)^s$. All the remaining terms of the formula depend on π only through the kinematic (as τ_M^S, τ_G^M) or geometry (as θ_M^G, D_M^S, D_G^M) of the rays (the values of the propagator at sources and receivers is supposed known).

B The gradient of the Born + Rays data misfit function

Let $J(\pi, r)$ be the data misfit function defined by (1) (37) (32). It is quadratical with respect to r, so the computation of its gradient is easy. Differentiating $J(\pi, r)$ with respect to r gives:

$$\delta J = -\sum_S \sum_G \int_0^T (d^{S,G}(t) - c_{\pi,r}^{S,G}(t)) \delta c^{S,G}(t) dt \tag{83}$$

If we define:

$$\begin{aligned} &e = d - c = \text{ residual section } = \\ &\text{part of the data which is not explained by the reflectivity } r \end{aligned} \tag{84}$$

and use the definition (37) (32) of c_G, formula (83) rewrites as:

$$\delta J = -\sum_S \sum_G \sum_M (\pi_M)^s A_M^{S,G} \delta r_M \int_0^T e^{S,G}(t) W_G(t) h * S^{(2+s)}(t - \tau_M^{S,G}) dt \tag{85}$$

Picking up the coefficient of δr_M yields $\partial J / \partial r_M(\pi, r) = (\nabla_r J(\pi, r))_M$:

$$[-\nabla_r J(\pi, r)]_M = (\pi_M)^s \sum_S \sum_G A_M^{S,G} \int_0^T e^{S,G}(t) W_G(t) h * S^{(2+s)}(t - \tau_M^{S,G}) dt \tag{86}$$

When $r = 0$, then $c = 0$ and the residual e coincides with the data d, so that:

$$[-\nabla_r J(\pi, 0)]_M = (\pi_M)^s \sum_S \sum_G A_M^{S,G} \int_0^T d^{S,G}(t) W_G(t) h * S^{(2+s)}(t - \tau_M^{S,G}) dt \tag{87}$$

But using the hypothesis (45) we can extend the integral to $(-\infty, +\infty)$ and use convolution notations:

$$[-\nabla_r J(\pi, 0)]_M = (\pi_M)^s \sum_S \sum_G A_M^{S,G} (W_G d_{S,G}) * h * S^{(2+s)}(\tau_M^{S,G}) \tag{88}$$

which can be rewritten as:

$$[-\nabla_r J(\pi, 0)]_M = (\pi_M)^s \sum_S \sum_G A_M^{S,G} E^{S,G}(\tau_M^{S,G}) \tag{89}$$

where:

$$E^{S,G} = h * (W_G d^{S,G}) * S^{(3)} = h * ((W_G)^2 h) * d_{ta}^{S,G} * S^{(2+s)}. \tag{90}$$

Using the relation (34) between the propagating wavelet $S(t)$ and the source function $f(t)$, we can rewrite (90) as:

$$E^{S,G} = (f^{(1+s)} * h * [(W_G)^2 h * d_{ta}^{S,G}])^{(s)} \tag{91}$$

We call $E^{S,G}$ the migration-filtered data, by opposition to $d^{S,G} = W_G h * d_{ta}^{S,G}$ which we called the misfit-filtered data.

Under the hypothesis (46) that the weights $W_G(t)$ vary slowly with respect to the signal $d_{ta}^{S,G}$, one can approximate $E^{S,G}(t)$ by:

$$E^{S,G} = (W_G)^2 h * h * S^{(2+s)} * d_{ta}^{S,G} \tag{92}$$

or equivalently:

$$E_G = (W_G)^2 h * h * f^{(1+s)} * d_{ta}^{S,G^{(s)}} \tag{93}$$

which ends the proof of formula (47) (48) (49) of the paper.

C Computing the best diagonal left inverse

Let A be a $n \times n$ matrix. We search for a diagonal matrix K such that

$$\|K^* A - I\|_\infty \leq \|KA - I\|_\infty \text{ for all diagonal matrices } K \tag{94}$$

But the matrix norm $\| \ \|_\infty$ defined by (19) is known to be given by:

$$\|KA - I\|_\infty = \max_{i=1\ldots n} \sum_{j=1\ldots n} |k_i a_{ij} - \delta_{ij}|$$

i.e.

$$\|KA - I\|_\infty = \max_{i=1\ldots n} \{|k_i| \sum_{j \neq i} |a_{ij}| + |k_i a_{ii} - 1|\} \tag{95}$$

The minimum of $\|KA - I\|_\infty$ will be obtained by chosing, for each $i = 1...n, k_i^* \in \mathbb{R}$ which minimizes $|k_i| \sum_{j\neq i} |a_{ij}| + |k_i a_{ii} - 1|$. As it can be seen on figure C-1, the solution is given by:

$$\begin{cases} k_i^* = 1/a_{ii} & \text{if} \quad |a_{ii}| > \sum_{j\neq i} |a_{ij}| \\ k_i^* = 0 & \text{if} \quad |a_{ii}| < \sum_{j\neq i} |a_{ij}|, \end{cases} \tag{96}$$

and the corresponding value of $\|KA - I\|_\infty$ is given by:

$$\|K^* A - I\|_\infty = \min\{1, \max_{i \ s.t. |a_{ii}| > \sum_{j\neq i} |a_{ij}|} \sum_{j \neq i} |a_{ij}|/|a_{ii}|\} \tag{97}$$

which satisfies:

$$0 \leq \|K^* A - I\|_\infty \leq 1. \tag{98}$$

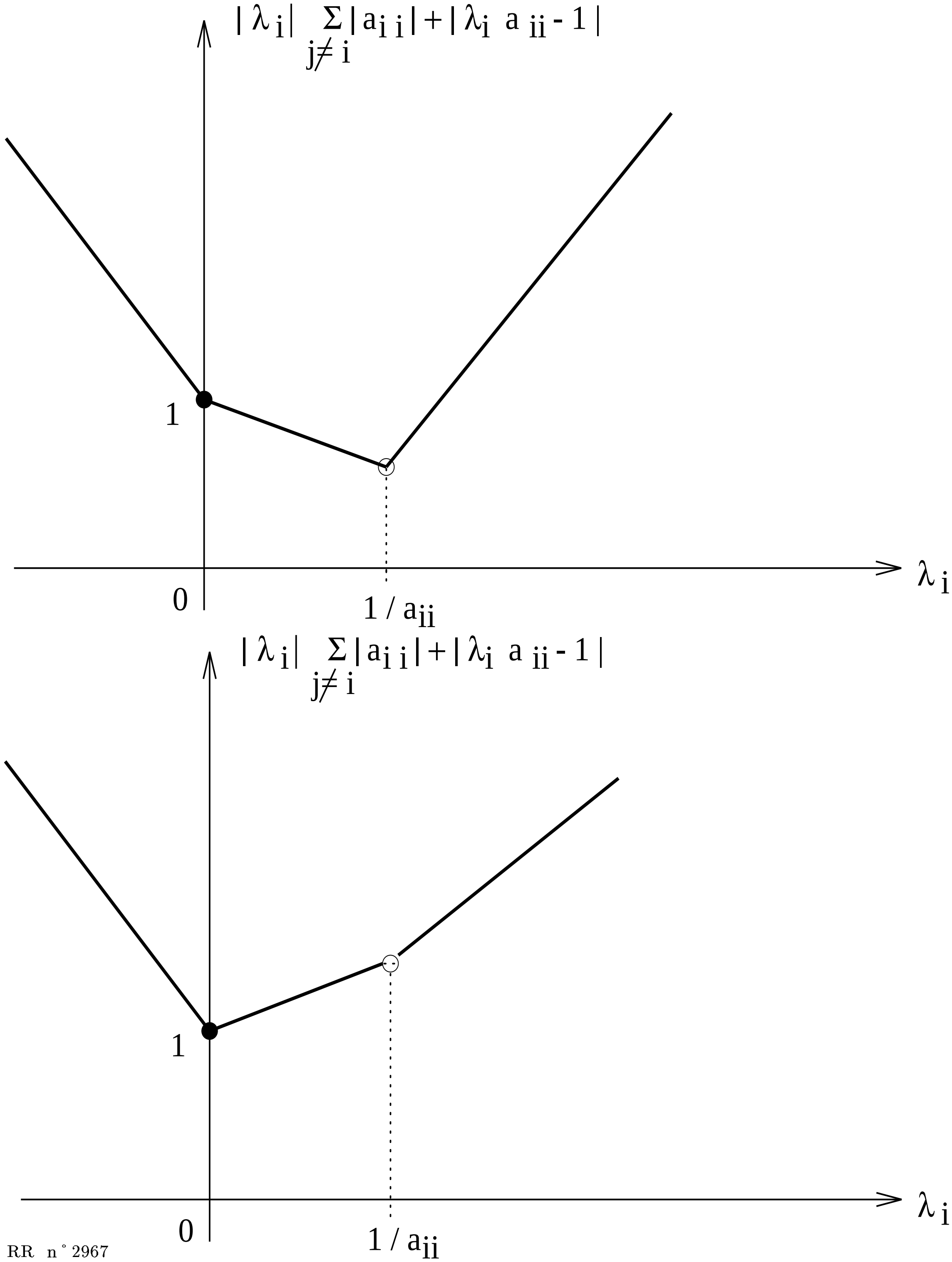

Figure 11: Graphical resolution of the minimization of $\|KA - I\|_\infty$

References

[1] A. Bamberger, G. Chavent, P. Lailly, and C. Hemon. Inversion of normal incidence seismograms. *Geophysics*, 47:737–770, 1982.

[2] G. Beylkin. The Inversion Problem and Applications of the Generalized Radon Transform. *Comm. Pure and Applied Math.*, XXXVII:579–599, 1984.

[3] G. Beylkin. Imaging of Discontinuities in the Inverse Scattering Problem by Inversion of a Causal Generalized Radon Transform. *J. Math. Phys*, 26-1, 1985.

[4] G. Beylkin and R. Burridge. Linearized inverse scattering problems in acoustics and elasticity. *Wave Motion*, 12:15–52, 1990.

[5] N. Bleistein. On the Imaging of Reflectors in the Earth. *Geophysics*, 52-7:931–942, July 1987.

[6] J. A. Carter and L. N. Frazer. Accomodating Lateral Velocity Changes in Kirchhoff Migration by Means of Fermat's Principle. *Geophysics*, 49-1:46–53, January 1984.

[7] G. Chavent and F. Clément. Separating propagation and reflection parameters in the acoustic wave equation. *Submitted to Geophysics*, 1994.

[8] P. Docherty. A Brief Comparison of Some Kirchhoff Integral Formula for Migration and Inversion. *Geophysics*, 56-8:1164–1169, August 1991.

[9] S. H. Gray, J. K. Cohen, and N. Bleisten. Velocity Inversion in a Stratified Medium with Separated Source and Receiver. *J. Acoustical Soc. Am.*, 68-1:234–240, July 1980.

[10] C. W. Groetch. The Theory of Tikhonov Regularization for Fredholm Equations of the First Kind. *Pitman*, 1984. Boston.

[11] T. H. Keho and W. B. Beydoun. Paraxial Ray Kirchhoff Migration. *Geophysics*, 53:1540–1546, 1988.

[12] P. Lailly. The seismic inverse problem as a sequence of before stack migrations. In J. Bee Bednar, Richard Redner, Enders Robinson, and Arthur Wegleing, editors, *Conference on Inverse Scattering : Theory and Application*, pages 206–220. SIAM, 1983.

[13] G. Lambaré. *Inversion linéarisée de données de sismique réflection par une méthode quasi-Newtonnienne*. PhD thesis, Université Paris VII, 27 Mai 1991. Thèse, Spécialité Sciences de la Thèse.

[14] A. Tarantola. Iinversion of seismic reflection data in the acoustic approximation. *Geophysics*, 49-8:1259–1266, August 1984.

[15] A. Tarantola. Linearized inversion of Seismic Reflection Data. *Geophysical prospecting*, 32:998–1005, 1984.

Unit´e de recherche INRIA Lorraine, Technopôle de Nancy-Brabois, Campus scientifique,
615 rue du Jardin Botanique, BP 101, 54600 VILLERS LES NANCY
Unit´e de recherche INRIA Rennes, Irisa, Campus universitaire de Beaulieu, 35042 RENNES Cedex
Unit´e de recherche INRIA Rhône-Alpes, 655, avenue de l' Europe, 38330 MONTBONNOT ST MARTIN
Unit´e de recherche INRIA Rocquencourt, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex
Unit´e de recherche INRIA Sophia-Antipolis, 2004 route des Lucioles, BP 93, 06902 SOPHIA-ANTIPOLIS Cedex

Éditeur
INRIA, Domaine de Voluceau, Rocquencourt, BP 105, 78153 LE CHESNAY Cedex (France)
ISSN 0249-6399